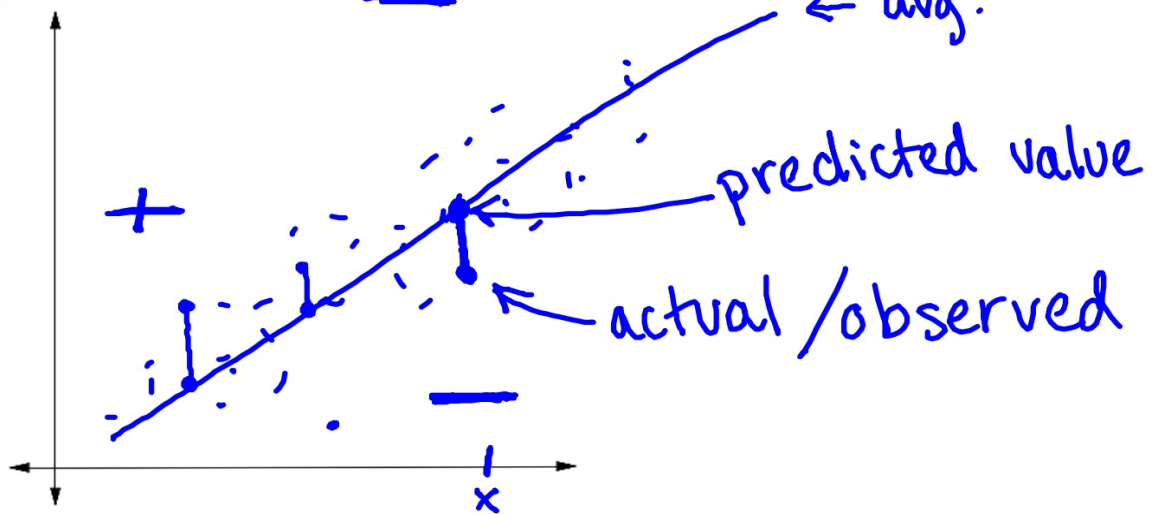


Chapter 8:

Line of Best Fit



Errors = residuals =

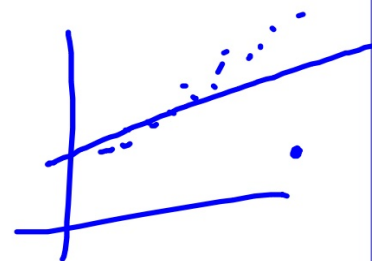
error in y-direction
actual - predicted (y)

Linear Regression Line:

- Straight line
- Describes how...
response variable (y) changes due
to explanatory variable (x)
- Used to ...
prediction!
- Requires that... clear explan. & resp.

Most accurate Regression line:

- Called: Least Squares Regression Line (LSRL)
- Definition: minimizes... errors (+ and -) in y-direction
- Form: $\hat{y} = b_0 + b_1 x$ $\hat{y} = a + bx$
hat ↑ intercept slope
- always ... passes thru (\bar{x}, \bar{y})
- not resistant to outliers
- on calculator:



$$\hat{y} = b_0 + b_1 x$$

$$\textcircled{2} = \bar{y} - b_1 \bar{x}$$

$$\textcircled{1} = r \left(\frac{s_y}{s_x} \right)$$

\bar{x}	$=$
s_x	$=$
\bar{y}	$=$
s_y	$=$
r	$=$

$$\hat{y} = b_0 + b_1 x$$

$$\hat{y} = 24.86 + 0.67(75)$$

$$Y1(75)$$

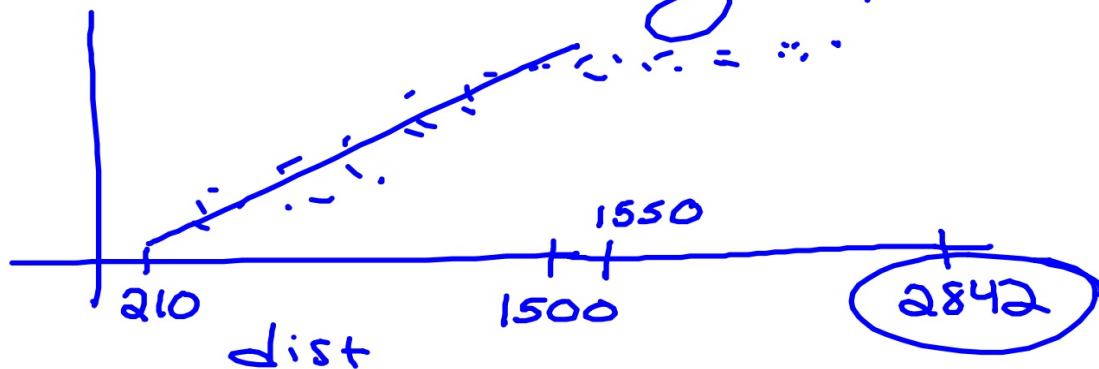
Complete worksheet 8A

$$b_1 = r \left(\frac{s_y}{s_x} \right) = 0.795 \left(\frac{59.45}{402.69} \right) \\ = 0.1174$$

$$b_0 = \bar{y} - b_1 \bar{x} = 166.92 - (0.1174)(712.67) \\ = 83.25$$

$$\hat{y} = 83.25 + 0.1174x$$

Residual = actual y - predicted y



$$\text{slope} = \frac{\Delta y}{\Delta x} = \frac{\$0.1174}{1 \text{ miles}}$$

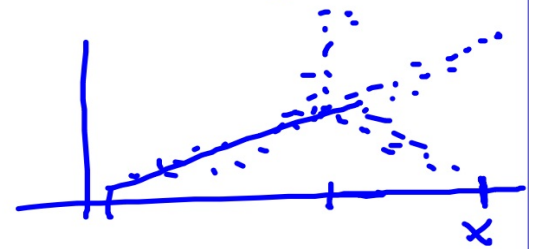
For every 1 mile, the airfare increases by \$0.12.

Vocab:

Extrapolation

bad!

predicting a y-value using an x that is far outside your given range.



Coefficient of determination:

symbol & calculation: r^2 $0 \leq r^2 \leq 1$ %

sentence interpretation: 63.2% of the change in air y-variable that is due to

Example:

$r = 0.80$ change in x-var. dist.

$r^2 = 0.64 \rightarrow 64\%$

Do the residuals example (#1) on the next page

$$\hat{y} = 17.886 - 0.587x$$

$$r = -0.759$$

$$r^2 = 0.576$$

Residuals (errors):

$$= \text{actual } y - \text{predicted } y$$

$$= y_i - \hat{y}_i$$

$$\text{sum of residuals} = 0$$

$$\text{mean of residuals} = 0$$

Residual Plot

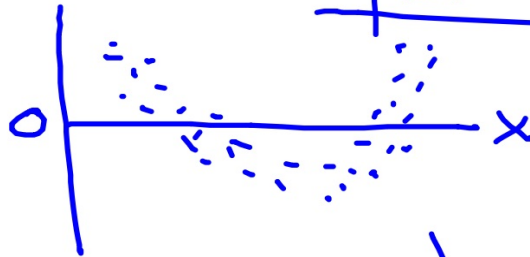
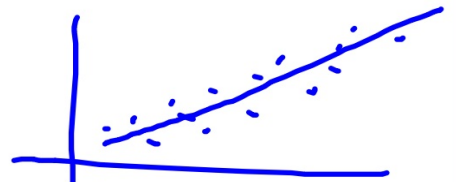
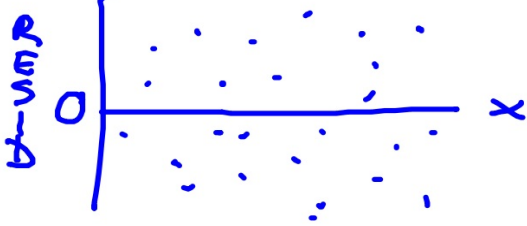
* Definition:

graph x -values vs. residuals
(y)

* Helps...

assess the fit of the LSRL

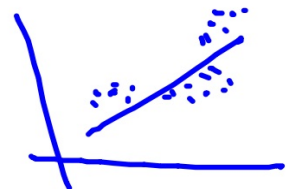
Examples:



No pattern = LSRL is a
good fit ~~for~~ for
data

pattern =

LSRL is
not the
best fit for
data



* Can show lurking variables

* on calculator:

Do examples #2 & 3 in the notes

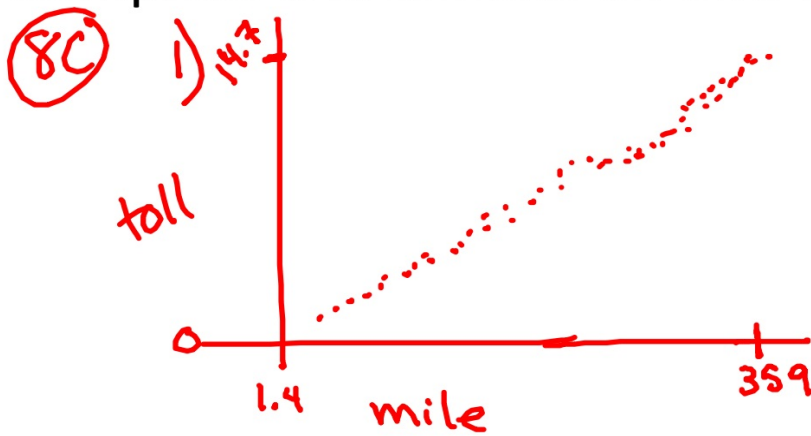
HW: finish Ex. 3
Wksht 8B #1-3

~~Computer Science 101~~

wiki! print! Ch. 8
classwork?

.

Complete worksheets 8B and 8C

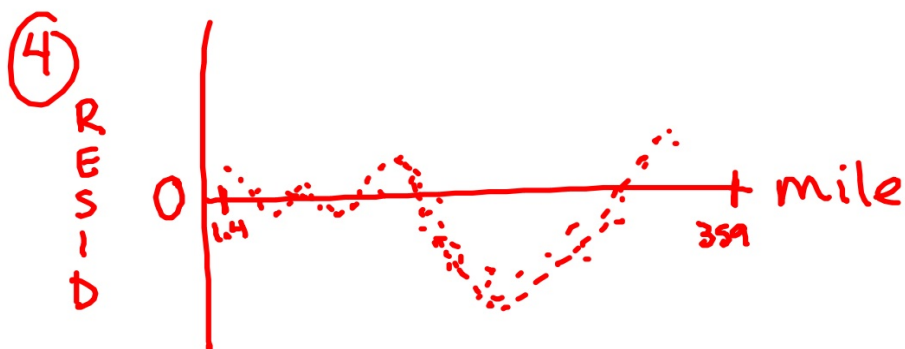


- linear
- very strong
- positive

②

$$\hat{y} = -0.1157 + 0.0401x$$
$$r = 0.999$$
$$r^2 = 0.9983$$

③ yes



⑤ no-linear model is not the best model b/c there's a pattern in the residual plot



⑦ For every 1 mile driven on the PA turnpike, the toll increases by $\sim 4\phi$.

⑧ 99.83% of the change in the toll prices is due to the miles driven.

Computer Outputs for Linear Regression:

An insurance company conducts a survey of 15 of its life insurance agents. The average number of minutes spent with each potential customer and the number of policies sold in a week are noted for each agent.

The following is a printout from the statistical analysis tool on Microsoft Excel.

Regression Statistics				
Multiple R	0.883620846			
R Square	0.780785799			
Adjusted R Square	0.783923168			
Standard Error	1.311483261			
Observations	15			

	Coefficients	Standard Error	t Stat	P-value
Intercept	-1.73106061	2.11612023	-0.81602	0.4128433
Minutes	0.549242424	0.080716215	6.80481	1.25E-05

- What is the equation of the LSR line relating minutes spent and policies sold?
 $\hat{y} = a + bx$ $\hat{y} = -1.731 + 0.549x$
- What is the value of r ? What is the value of r^2 ?
 $r = 0.8836$ $r^2 = 0.781$
- Interpret the slope in the context of the problem

The following is a MINITAB regression printout relating average number of degree-days per month to gas consumption (in cubic feet).

Predictor	Coef	StDev	T	P
Constant	123.24	28.60	4.31	0.004
Degree-d	20.221	1.145	17.66	0.000
S=43.45				
R-sq=97.8%				
R-sq(adj)=97.5%				

- What is the equation of the LSR line relating degree days to gas consumption?

$$\hat{y} = 123.24 + 20.221x$$

- What is the value of r ? What is the value of r^2 ?

$$r = 0.989 \quad r^2 = 97.8\%$$

- Interpret the slope in the context of the problem?

$$\frac{\Delta y}{\Delta x} = \frac{20.221 \text{ ft}^3 \text{ of gas}}{1 \text{ degree day}}$$

For every 1 degree day, the gas consumption increases on avg. by 20.221 ft^3 .

Response attribute (numeric): MPG

Predictor	Coefficient	Std Error	t Statistic	P Value	ΔR ²
Constant	62.5416	2.7885	22.428	0.0000	
Weight	-0.0109	0.0009	-12.741	0.0000	0.9206

Source	Degrees of Freedom	Sum of Squares	Mean Square	F Statistic	P Value	ΔR ²
Regression	1	620.431	620.431	162.337	0.0000	0.9206
Residual	14	53.506	3.822			
Total	15	673.938				

R-Squared: 0.920607

Adjusted R-Squared: 0.914936

Standard Deviation of the Error: 1.95496

$$\hat{y} = 62.5416 - 0.0109x$$

$$r^2 = 0.9206 \quad r = -0.95$$

$$\hat{y} = b_0 + b_1 x$$

$$b_1 = r \left(\frac{s_y}{s_x} \right)$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

#4

$$z = 1.5$$

$$1.5 = \frac{x - \bar{x}}{s_x}$$

wine

Do LSRL worksheet